



Penerapan Algoritma C4.5 Untuk Prediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan (SPP) Santri

Bayu Angkoso¹, Irmayansyah^{2*}

¹Sistem Informasi/Universitas Binaniga Indonesia

Email: bayuangkoso@gmail.com

²Sistem Informasi/Universitas Binaniga Indonesia

Email: irma@unbin.ac.id

*) *Corresponding Author*

ABSTRACT

SPP is a mandatory fee for students used by Islamic boarding schools to facilitate all learning activities carried out by students, with a predetermined payment time. One of the problems in paying tuition fees is the case of students being late in paying tuition fees. The problem of late payment of tuition fees is something that must be considered because this can affect the implementation of education and fulfillment of needs and result in a decrease in the financial income of the pesantren. If the problem of late tuition payments can be predicted more quickly, then the management can prevent and anticipate earlier. To overcome this problem, the c4.5 algorithm is applied to predict late payment of tuition fees in order to produce a pattern based on the classification results. By using the variables of Parents' Income, Dependents, Father's Job, First Payment, Second Payment and Third Payment. This is done to see students who have the potential to be late in making tuition payments so that they can anticipate a decline in the financial income of the pesantren. In this study, the accuracy of the c.45 algorithm method has been tested for prediction of late tuition payments using the confusion matrix formula with an accuracy of 76.92%.

Keywords: Prediction; Confussion Matrix; payment; accuracy; algorithm.

ABSTRAK

SPP merupakan iuran wajib bagi santri yang dipergunakan oleh pondok pesantren untuk memfasilitasi segala kegiatan pembelajaran yang dilakukan santri, dengan waktu pembayaran ditentukan sebelumnya. Salah satu masalah dalam pembayaran SPP adalah kasus santri terlambat dalam melakukan pembayaran SPP. Permasalahan keterlambatan Pembayaran SPP menjadi sesuatu yang harus di perhatikan karena hal ini dapat mempengaruhi pelaksanaan Pendidikan dan pemenuhan kebutuhan serta mengakibatkan penurunan pendapatan keuangan pesantren. Jika masalah keterlambatan pembayaran SPP dapat di prediksi lebih cepat, maka pihak manajemen dapat mencegah dan mengantisipasi lebih awal. Untuk mengatasi masalah tersebut maka diterapkan algoritma c4.5 untuk prediksi keterlambatan pembayaran SPP agar menghasilkan pola berdasarkan dengan hasil klasifikasi. Dengan menggunakan variabel Penghasilan Orang Tua, Tanggungan, Pekerjaan Ayah, Pembayaran Pertama, Pembayaran Kedua dan Pembayaran Ketiga. Hal ini dilakukan untuk melihat santri yang berpotensi telat melakukan pembayaran SPP sehingga dapat mengantisipasi penurunan pendapatan keuangan pesantren. pada penelitian ini sudah dilakukan uji akurasi metode algoritma c.45 untuk prediksi keterlambatan pembayaran SPP dengan menggunakan rumus *confussion matrix* dengan hasil akurasi 76.92%.

Keywords: Prediksi; Confussion Matrix; Pembayaran; Akurasi; Algoritma.

A. PENDAHULUAN

Sumbangan Pembinaan Pendidikan atau biasa disingkat SPP merupakan iuran rutin yang harus dibayarkan santri setiap satu bulan sekali. Iuran rutin atau SPP ini merupakan kewajiban bagi santri yang masih aktif di pondok pesantren. Tujuan dari SPP atau Sumbangan Pembinaan Pendidikan adalah untuk biaya operasional penyelenggaraan pendidikan, kesejahteraan karyawan, perbaikan sarana dan tentunya membantu pembinaan pendidikan. Oleh karena itu sumbangan Pembinaan Pendidikan (SPP) memegang peranan penting dalam pembiayaan operasional di pesantren.

Banyak permasalahan terkait pembiayaan operasional pondok diantaranya kasus santri terlambat dalam melakukan pembayaran SPP meskipun sudah ditetapkan batas waktu pembayarannya. Oleh karena itu, perlu dilakukan pendeteksian faktor-faktor yang menjadi penyebab terjadinya keterlambatan pembayaran SPP dengan menggunakan data mining.

Data Mining telah banyak diterapkan sebagai solusi terhadap permasalahan pada dunia nyata dalam bidang keilmuan dan bisnis. *Data mining* merupakan suatu aktifitas eksplorasi dan analisis, dari sejumlah besar data untuk menemukan pola-pola dan aturan-aturan yang berguna. Metode – metode yang terdapat dalam *data mining* yaitu Algoritma klasifikasi yang populer diantaranya *ada Decision Tree*, *Naïve Bayes* dan *Neural Network*. Algoritma *Decision Tree* bekerja dengan membentuk pohon keputusan yang dapat disimpulkan aturan – aturan klasifikasi tertentu, salah satunya C4.5. Algoritma *Decision Tree* ini memiliki beberapa kelebihan diantaranya mudah untuk diinterpretasikan, dapat menggunakan data numerik dan kategorikal, tidak membutuhkan biaya yang mahal saat membangun algoritma ini, stabil dan cepat bila digunakan dalam dataset yang besar, proses pengambilan keputusan dapat dipahami dengan mudah (F. Gorune scu, 2011).

Algoritma C4.5 termasuk ke dalam teknik data mining, dan data mining memiliki beberapa tahapan yaitu: pembersihan data (*cleaning data*), integrasi data, seleksi data, transformasi data yaitu pengubahan data menjadi format ekstensi yang sesuai untuk pengolahan dalam data mining, kemudian memproses data mining, lalu mengevaluasi pola atau mengidentifikasi pola-pola menarik kedalam *knowledge-based* yang diidentifikasi. Kemudian merepresentasikan pengetahuan yaitu visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atas aksi dari hasil analisis yang di dapat. Algoritma C4.5 merupakan struktur pohon dimana terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Algoritma C4.5 menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal.

Untuk mengatasi masalah keterlambatan pembayaran SPP maka diterapkan algoritma c4.5 untuk prediksi keterlambatan pembayaran spp agar menghasilkan pola berdasarkan dengan hasil klasifikasi. Dengan menggunakan variabel Penghasilan Orang Tua, Tanggungan, Pekerjaan Ayah, Pembayaran Pertama, Pembayaran Kedua dan Pembayaran Ketiga. Hal ini dilakukan untuk melihat santri yang berpotensi telat melakukan pembayaran SPP sehingga dapat mengantisipasi penurunan pendapatan keuangan pesantren

B. METODE

1. Algoritma C4.5

Metode konseptual yang digunakan dalam penelitian ini yaitu metode Algoritma C4.5. Menurut (Larose, 2005) Metode Algoritma C4.5 merupakan metode yang banyak digunakan untuk klasifikasi data yang memiliki atribut numerik dan kategorikal. Hasil proses klasifikasi berupa rules yang dapat dipergunakan untuk memprediksi nilai atribut tipe diskrit record baru. Tahapan yang dilakukan untuk memprediksi keterlambatan pembayaran SPP dengan menerapkan metode algoritma C4.5 adalah :

a. Pengumpulan Data

- b. Pemilihan Data (Data Selection)
- c. Penentuan Variabel
- d. Memilih atribut menjadi akar.
- e. Membuat cabang untuk disetiap nilai.
- f. Membagi kasus dalam setiap cabang.
- g. Mengulangi proses pada point d, e dan f, untuk setiap cabang hingga semua kasus pada cabang tersebut memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada. Untuk menghitung Gain digunakan rumus seperti tertera dalam persamaan (1) berikut.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i)$$

Keterangan:

- S : himpunan kasus
- A : atribut
- n : jumlah partisi atribut A
- |S_i| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S

Sementara itu perhitungan nilai entropy dapat dilihat pada persamaan (2) berikut.

$$\text{Entropy}(S) = \sum_{i=1}^n - P_i * \text{Log}_2 p_i$$

Keterangan:

- S : himpunan kasus
- A : fitur
- n : jumlah partisi S
- P_i : proporsi dari S_i terhadap S

Proses pengulangan pada metode decision tree ini akan berhenti apabila :

- a. Semua data telah terbagi rata
- b. Tidak ada lagi atribut yang bisa dibagi lagi
- c. Tidak ada data record dalam cabang yang kosong

2. Teknik Analisa Data

Untuk menguji akurasi penerapan metode algoritma C.45 pada prediksi keterlambatan pembayaran SPP digunakan confusion matrix. *Confussion Matrix* adalah metode untuk mengevaluasi model klasifikasi dengan memperkirakan objek benar atau salah (F. Gorunescu, 2011). Matriks tersebut membandingkan hasil prediksi dengan informasi nilai aktual dalam klasifikasi

Tabel 1. Confusion Matrix

Classification	Predicted class	
	Class = Yes	Class = No
Class = Yes	a (<i>true positive-TP</i>)	b (<i>false negative-FN</i>)
Class = No	c (<i>false positive-FP</i>)	d(<i>true negative-TN</i>)

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{A + D}{A + B + C + D}$$

Keterangan:

- A = jika hasil prediksi positif dan data sebenarnya positif
- B = jika hasil prediksi negatif dan data sebenarnya positif
- C = jika hasil prediksi positif dan data sebenarnya negative

D = jika hasil prediksi negatif dan data sebenarnya negatif

C. HASIL DAN PEMBAHASAN

1. Hasil

a. Data Selection

Tabel 2 menunjukkan hasil proses data selection. Seleksi yang dilakukan adalah dengan menghapus filed – filed yang tidak digunakan, menghapus isi filed yang kosong pada data santri tahun akademik 2019/2020, dan hanya memilih field dan isi filed yang digunakan sebagai variabel seperti Penghasilan Orang Tua, Tanggungan, Pekerjaan Ayah, Pembayaran Pertama, Pembayaran Kedua dan Pembayaran Ketiga.

Tabel 2. Data Selection

No	Nama	Penghasilan Orang Tua	Tanggungan	Pekerjaan Ayah	P1	P2	P3	Status
1	ADITYA RIYADI	4.000.000	2	Wiraswasta	9-Jul-19	1-Sep-19	6-Okt-19	Telat
2	EDITYA RIYADI	4.000.000	2	Wiraswasta	3-Jul-19	1-Agu-19	6-Okt-19	Telat
3	FAHREZA NUR QOLBI SALIM	4.500.000	2	Wiraswasta	7-Jul-19	9-Aug-29	29-Sep-19	Tepat
4	HADDY NUR ALIM	4.000.000	2	Karyawan Swasta	10-Jul-19	7-Sep-19	7-Sep-19	Tepat
5	MARSEL RAMADANI	5.000.000	3	Wiraswasta	5-Agu-19	19-Nov-19	2-Des-19	Telat
6	RAJA HAIKAL MAHFUDZ	4.000.000	2	Wiraswasta	30-Jun-19	11-Agu-19	6-Des-19	Telat
..
43	RAVIYAN	4.000.000	3	Karyawan Swasta	21-Jul-19	4-Agu-19	1-Sep-19	Tepat

b. Data Transformation

Tahap selanjutnya yaitu melakukan proses transformasi dengan cara mengubah pengetahuan dari data asli (real) yang terdapat pada tabel 2. Transformasi data dilakukan pada kriteria tanggungan dan pembayaran yang dilakukan dengan cara mengelompokkan berdasarkan pada batas minimum dan maksimum. berikut merupakan hasil transformasi kedua kriteria tersebut.

Tabel 3. Kriteria Tanggungan

Rata-rata Tanggungan	Kriteria
1 Anak	Sedikit
2 – 3 Anak	Cukup
>4 Anak	Banyak

Tabel 4. Kriteria Pembayaran

Rata-rata SKS	Kriteria
< Tanggal 10	Tepat
>Tanggal 10	Telat

Setelah dilakukan transformasi kepada kedua kriteria tersebut, langkah selanjutnya adalah mempersiapkan data hasil seleksi yang sudah di transformasi. Tabel 5 menunjukkan data hasil seleksi yang sudah di transformasi

Tabel 5. Data Seleksi setelah ditransformasi

No	Nama	Penghasilan Orang Tua	Tanggungan	Pekerjaan Ayah	P1	P2	P3	Status
1	Ahmad Roi Berliansyah	> 4 Juta	Banyak	Wiraswasta	Tidak	Tidak	Tidak	Telat
2	Daffa Khoirul Ikhwan	> 4 Juta	Cukup	Karyawan Swasta	Ya	Tidak	Ya	Tepat
3	Elan Albadri	> 4 Juta	Cukup	Wiraswasta	Tidak	Tidak	Ya	Telat
4	M Fahri	> 4 Juta	Cukup	Karyawan Swasta	Tidak	Ya	Ya	Telat
5	Mohammad Amar Faros	> 4 Juta	Cukup	Lainnya	Tidak	Ya	Ya	Tepat
6	Muhammad Nurhidayat	> 4 Juta	Cukup	Karyawan Swasta	Tidak	Tidak	Ya	Telat
...
43	Gadis Nuraini	2 -4 Juta	Cukup	Wiraswasta	Tidak	Ya	Tidak	Tepat

c. Penentuan Variabel dalam Penelitian

Variabel yang digunakan dalam prediksi keterlambatan pembayaran SPP santri adalah Penghasilan Orangtua, Tanggungan, Pekerjaan Ayah, Pembayaran Bulanan Pertama, Pembayaran Bulanan Kedua dan Pembayaran Bulanan Ketiga.

1) Penghasilan Orang Tua

Variabel Penghasilan Orang Tua seperti yang ada pada tabel 6

Tabel 6. Penghasilan Orang Tua

Variable	Nilai
Penghasilan Orang tua	Kurang Dari 1 Juta
	1 Sampai Dengan 2 Juta
	2 Sampai Dengan 4 Juta
	Lebih Dari 4 Juta

2) Tanggungan

Variabel Tanggungan seperti yang ada pada tabel 7

Tabel 7. Tanggungan

Variable	Nilai
Tanggungan	Sedikit
	Cukup
	Banyak

3) Pekerjaan Ayah

Variabel Pekerjaan Ayah seperti yang ada pada tabel 8

Tabel 8 Variabel Pekerjaan Ayah

Variable	Nilai
Pekerjaan Ayah	Wiraswasta
	Lainnya
	PNS
	Karyawan Swasta
	Buruh

4) Pembayaran Bulanan Pertama

Variabel Pembayaran Bulanan Pertama seperti yang ada pada tabel 9

Tabel 9. Variabel Pembayaran Bulanan Pertama

Variable	Nilai
Pembayaran Bulanan Pertama	Ya
	Tidak

5) Pembayaran Bulanan Kedua

Variabel Pembayaran Bulanan Kedua seperti yang ada pada tabel 10

Tabel 10. Variabel Pembayaran Bulanan Kedua

Variable	Nilai
Pembayaran Bulanan Kedua	Ya
	Tidak

6) Pembayaran Bulanan Ketiga

Variabel Pembayaran Bulanan Ketiga seperti yang ada pada tabel 11.

Tabel 11. Variabel Pembayaran Bulanan ketiga

Variable	Nilai
Pembayaran Bulanan Ketiga	Ya
	Tidak

d. Menghitung Gain dan Entrophy

Gunakan jumlah kasus yang ada, jumlah kasus hasil pengurangan, dan jumlah kasus hasil penjumlahan untuk perhitungan. Kemudian melakukan perhitungan untuk mencari nilai Entrophy dari setiap atribut dan nilai Gain dari setiap atribut.

$$\text{Jumlah kejadian} = 43$$

$$\text{Tepat} = 24$$

$$\text{Telat} = 19$$

1) Entropy

Entrophy (Total)

$$= \left(-\frac{24}{43} \times \log_2 \left(\frac{24}{43}\right)\right) + \left(-\frac{19}{43} \times \log_2 \left(\frac{19}{43}\right)\right) = 0,9902$$

Entrophy (Penghasilan)

Entrophy (<1 Juta)

$$= \left(-\frac{0}{0} \times \log_2 \left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} \times \log_2 \left(\frac{0}{0}\right)\right) = 0$$

Entrophy (1-2 Juta)

$$= \left(-\frac{2}{3} \times \log_2 \left(\frac{2}{3}\right)\right) + \left(-\frac{1}{3} \times \log_2 \left(\frac{1}{3}\right)\right) = 0,9183$$

Entrophy (2-4 Juta)

$$= \left(-\frac{14}{21} \times \log_2 \left(\frac{14}{21}\right)\right) + \left(-\frac{7}{21} \times \log_2 \left(\frac{7}{21}\right)\right) = 0,9183$$

Entrophy (>4 Juta)

$$= \left(-\frac{8}{19} \times \log_2 \left(\frac{8}{19}\right)\right) + \left(-\frac{11}{19} \times \log_2 \left(\frac{11}{19}\right)\right) = 0,9819$$

Entrophy (Tanggungan)

Entrophy (Sedikit)

$$= \left(-\frac{2}{2} \times \log_2 \left(\frac{2}{2}\right)\right) + \left(-\frac{0}{2} \times \log_2 \left(\frac{0}{2}\right)\right) = 0$$

Entrophy (Cukup)

$$= \left(-\frac{20}{33} \times \log_2 \left(\frac{20}{33}\right)\right) + \left(-\frac{13}{33} \times \log_2 \left(\frac{13}{33}\right)\right) = 0,9673$$

Entrophy (Banyak)

$$= \left(-\frac{2}{8} \times \log_2 \left(\frac{2}{8}\right)\right) + \left(-\frac{6}{8} \times \log_2 \left(\frac{6}{8}\right)\right) = 0,8113$$

2) Gain

Setelah didapatkan nilai Entropy dari setiap atribut maka selanjutnya menghitung nilai Gain yaitu dengan cara Entropy total dikurangi jumlah hasil perkalian nilai dibagi jumlah total kasus dikali Entropy dari masing – masing kasus. Perhitungan masing – masing nilai Gain dari empat atribut adalah sebagai berikut:

Gain (Total, Penghasilan)

$$= 0,9902 - \left(\left(\frac{0}{43} \right) \times 0 \right) + \left(\left(\frac{3}{43} \right) \times 0,9183 \right) + \left(\left(\frac{21}{43} \right) \times 0,9183 \right) + \left(\left(\frac{19}{43} \right) \times 0,9819 \right) = 0,0438$$

Gain (Total, Tanggungan)

$$= 0,9902 - \left(\left(\frac{2}{43} \right) \times 0 \right) + \left(\left(\frac{33}{43} \right) \times 0,9673 \right) + \left(\left(\frac{8}{43} \right) \times 0,8113 \right) = 0,0969$$

Gain (Total, Pekerjaan Ayah)

$$= 0,9902 - \left(\left(\frac{16}{43} \right) \times 0,9544 \right) + \left(\left(\frac{20}{43} \right) \times 0,9341 \right) + \left(\left(\frac{3}{43} \right) \times 0,9183 \right) + \left(\left(\frac{3}{43} \right) \times 0,9183 \right) + \left(\left(\frac{1}{43} \right) \times 0 \right) = 0,0725$$

Gain (Total, P1)

$$= 0,9902 - \left(\left(\frac{17}{43} \right) \times 0,7871 \right) + \left(\left(\frac{26}{43} \right) \times 0,9829 \right) = 0,0847$$

Gain (Total, P2)

$$= 0,9902 - \left(\left(\frac{15}{43} \right) \times 0,9710 \right) + \left(\left(\frac{28}{43} \right) \times 0,9963 \right) = 0,0028$$

Gain(Total, P3)

$$= 0,9902 - \left(\left(\frac{22}{43} \right) \times 0,8454 \right) + \left(\left(\frac{21}{43} \right) \times 0,9587 \right) = 0,0895$$

3) Hasil yang di dapat dari perhitungan *Entropy* dan *Gain*

Tabel 12 Hasil Perhitungan Entropy dan Gain

		Jumlah	Tepat	Telat	Entropy	Gain
Total		43	24	19	0,9902	
Penghasilan						0,0438
	< 1 Juta	0	0	0	0	
	1 - 2 Juta	3	2	1	0,9183	
	2 -4 Juta	21	14	7	0,9183	
	> 4 Juta	19	8	11	0,9819	
Tanggungan						0,0969
	Sedikit	2	2	0	0	
	Cukup	33	20	13	0,9673	
	Banyak	8	2	6	0,8113	
Pekerjaan Ayah						0,0725
	WIRASWASTA	16	6	10	0,9544	
	KARYAWAN SWASTA	20	13	7	0,9341	
	BURUH	3	2	1	0,9183	
	PNS	3	2	1	0,9183	
	LAINNYA	1	1	0	0	
P1						0,0847
	Ya	17	13	4	0,7871	
	Tidak	26	11	15	0,9829	
P2						0,0028
	Ya	15	9	6	0,9710	
	Tidak	28	15	13	0,9963	
P3						0,0895
	Ya	22	16	6	0,8454	
	Tidak	21	8	13	0,9587	

4) Pohon Keputusan

Dari hasil perhitungan nilai *Entropy* dan *Gain* seperti terlihat pada tabel 12 dapat diketahui bahwa atribut yang memiliki nilai *Gain* tertinggi adalah atribut **Tanggung** yaitu sebesar **0,0969** dengan demikian, atribut **Tanggung** menjadi node akar. Ada tiga nilai yang terdapat pada atribut **Tanggung** yaitu Sedikit, Cukup, dan banyak. Dari ketiga nilai tersebut masih harus dilakukan perhitungan lebih lanjut dikarenakan belum memiliki nilai kelas yang sama. Gambar 1 menunjukkan pohon keputusan untuk atribut tanggung.



Gambar 1. Perhitungan

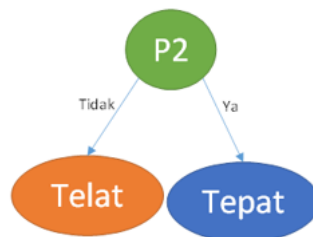
5) Perhitungan Node 1.2

Gunakan jumlah kasus yang ada, jumlah kasus hasil pengurangan, dan jumlah kasus hasil penjumlahan untuk perhitungan. Kemudian melakukan perhitungan untuk mencari nilai **Entropy** dari setiap atribut dan nilai **Gain** dari setiap atribut.

Tabel 13. Perhitungan Node 1.2

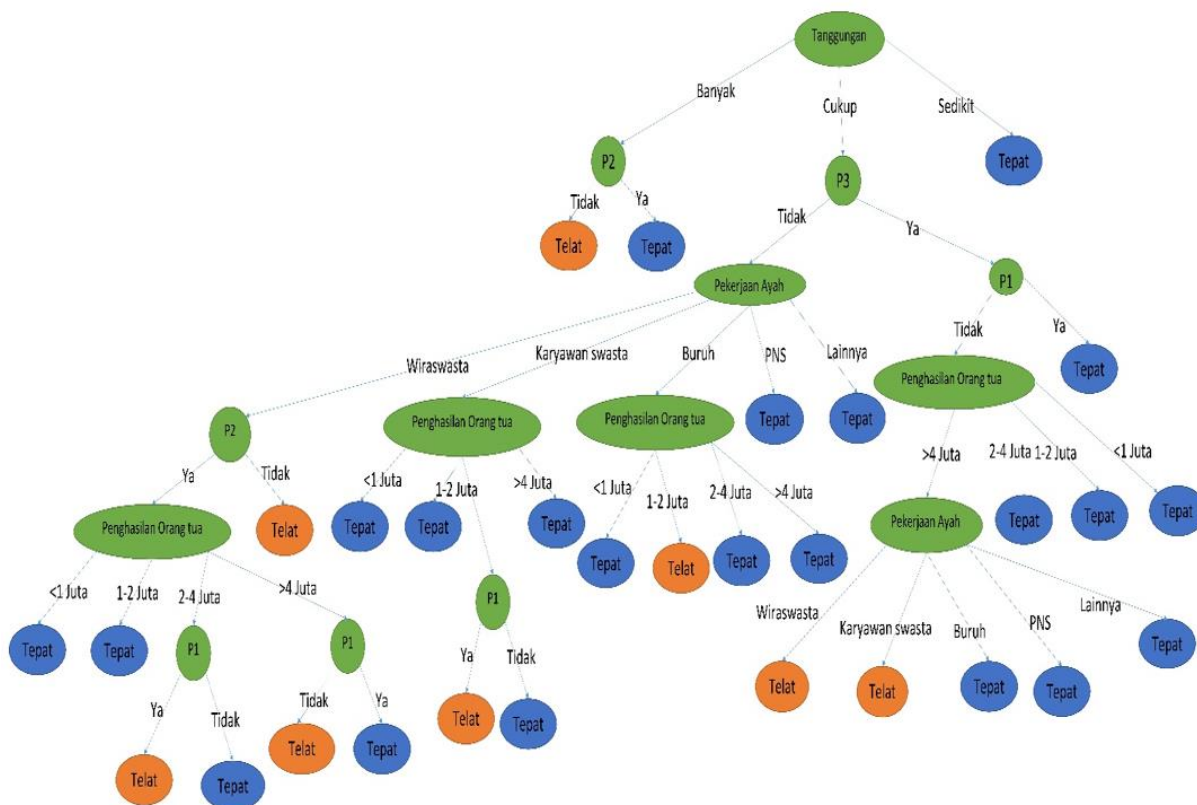
		Jumlah	Tepat	Telat	Entropy	Gain
Total		8	2	6	0,8113	
Penghasilan						0,4056
	< 1 Juta	0	0	0	0	
	1 - 2 Juta	1	1	0	0	
	2 -4 Juta	4	1	3	0,8113	
	> 4 Juta	3	0	3	0	
Pekerjaan Ayah						0,1556
	WIRASWASTA	4	1	3	0,8113	
	KARYAWAN SWASTA	2	0	2	0	
	BURUH	0	0	0	0	
	PNS	2	1	1	1	
	LAINNYA	0	0	0	0	
P1						0
	Ya	0	0	0	0	
	Tidak	8	2	6	0,8113	
P2						0,8113
	Ya	2	2	0	0	
	Tidak	6	0	6	0	
P3						0,0157
	Ya	3	1	2	0,9183	
	Tidak	5	1	4	0,7219	

Dari hasil perhitungan nilai *Entropy* dan *Gain* seperti terlihat pada tabel 12 dapat diketahui bahwa atribut yang memiliki nilai *Gain* tertinggi adalah atribut **P2** yaitu sebesar 1 dengan demikian, atribut **P2** menjadi node cabang 1.2 Ada dua nilai yang terdapat pada atribut yaitu **Ya dan Tidak**. Dari kedua nilai tersebut nilai **Ya** dinyatakan Tepat dan nilai **Tidak** dinyatakan Telat. Dari hasil tersebut dapat di gambarkan pohon keputusan (decision tree) dapat dilihat pada Gambar 2 .



Gambar 2. Pohon keputusan Node 1.2

- 6) Tampilan Pohon Keputusan Akhir
 Berdasarkan hasil perhitungan dengan menggunakan metode algoritma C4.5 maka diperoleh hasil berupa pohon keputusan (*decision tree*) seperti pada gambar 3 yang akan digunakan sebagai dasar untuk memprediksi keterlambatan pembayaran SPP santri



Gambar 3. Pohon Keputusan

e. Rule Hasil Dari Perhitungan Algoritma C 4.5

Dengan memperhatikan pohon keputusan pada Gambar 5 diketahui bahwa semua kasus sudah masuk dalam kelasnya masing-masing, dari pohon yang terbentuk dihasilkan sejumlah aturan. Adapun aturan yang terbentuk adalah berikut:

- Jika Tanggungan (Sedikit) = Tepat
- Jika Tanggungan (Cukup) & P3 (Ya) & P1 (Ya) = Tepat
- Jika Tanggungan (Cukup) & P3 (Ya) & P1 (Tidak) & penghasilan(< 1 Juta) = Tepat
- Jika Tanggungan (Cukup) & P3 (Ya) & P1 (Tidak) & penghasilan(1 Sampai 2 Juta) = Tepat
- Jika Tanggungan (Cukup) & P3 (Ya) & P1 (Tidak) & penghasilan(2 Sampai 4 Juta) = Tepat
- Jika Tanggungan (Cukup) & P3 (Ya) & P1 (Tidak) & penghasilan(Lebih Dari 4 Juta) & pekerjaan_ayah(WIRASWASTA) = Telat
- Jika Tanggungan (Cukup) & P3 (Ya) & P1 (Tidak) & penghasilan(Lebih Dari 4 Juta) & pekerjaan_ayah(KARYAWAN SWASTA) = Telat

Jika Tanggungan (Cukup) & P3 (Ya) & P1 (Tidak) & penghasilan(Lebih Dari 4 Juta) & pekerjaan_ayah(BURUH) = Tepat
Jika Tanggungan (Cukup) & P3 (Ya) & P1 (Tidak) & penghasilan(Lebih Dari 4 Juta) & pekerjaan_ayah(PNS) = Tepat
Jika Tanggungan (Cukup) & P3 (Ya) & P1 (Tidak) & penghasilan(Lebih Dari 4 Juta) & pekerjaan_ayah(LAINNYA) = Tepat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah(WIRASWASTA) & P2 (Ya) & penghasilan(< 1 Juta) = Telat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah(WIRASWASTA) & P2 (Ya) & penghasilan(1 Sampai 2 Juta) = Telat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah(WIRASWASTA) & P2 (Ya) & penghasilan(2 Sampai 4 Juta) & P1 (Ya) = Telat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah(WIRASWASTA) & P2 (Ya) & penghasilan(2 Sampai 4 Juta) & P1 (Tidak) = Tepat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah(WIRASWASTA) & P2 (Ya) & penghasilan(Lebih Dari 4 Juta) & P1 (Ya) = Tepat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (WIRASWASTA) & P2 (Ya) & penghasilan (Lebih Dari 4 Juta) & P1 (Tidak) = Telat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (WIRASWASTA) & P2 (Tidak) = Telat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (KARYAWAN SWASTA) & penghasilan (< 1 Juta) = Tepat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (KARYAWAN SWASTA) & penghasilan (1 Sampai 2 Juta) = Tepat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (KARYAWAN SWASTA) & penghasilan (2 Sampai 4 Juta) & P1 (Ya) = Telat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (KARYAWAN SWASTA) & penghasilan (2 Sampai 4 Juta) & P1 (Tidak) = Tepat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (KARYAWAN SWASTA) & penghasilan (Lebih Dari 4 Juta) = Telat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (BURUH) & penghasilan (< 1 Juta) = Tepat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (BURUH) & penghasilan (1 Sampai 2 Juta) = Telat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (BURUH) & penghasilan (2 Sampai 4 Juta) = Tepat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (BURUH) & penghasilan (Lebih Dari 4 Juta) = Tepat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (PNS) = Tepat
Jika Tanggungan (Cukup) & P3 (Tidak) & pekerjaan_ayah (LAINNYA) = Tepat
Jika Tanggungan (Banyak) & P2 (Ya) = Tepat
Jika Tanggungan (Banyak) & P2 (Tidak) = Telat
Hasil klasifikasi pada data sampel atribut Usia sebagai root pada decision tree, sedangkan atribut lainnya sebagai child node, dari data sampel dengan 43 record dihasilkan jumlah aturan yang terbentuk sebanyak 30 aturan.

2. Pembahasan

Pada tahap ini dilakukan pengukuran keakuratan antara hasil yang dicapai menggunakan confusion matrix. Pengukuran akurasi dilakukan dengan membandingkan hasil prediksi data latih berdasarkan variabel yang telah ditentukan dengan data yang seharusnya atau data nyata. Dari Hasil pengukuran tersebut diperoleh hasil akurasi sebesar 76,92%, yang diperoleh dari jawaban prediksi dan data sebenarnya positif (TP) ditambah jumlah jawaban prediksi negatif dan dengan sebenarnya negatif (TN) di bagi dengan pejumlahan prediksi positif dan data sebenarnya positif (TP) ditambah jawaban prediksi positif dan data sebenarnya negatif (FP) , ditambah jumlah jawaban prediksi negatif dan dengan sebenarnya negatif (TN), ditambah

jawaban prediksi negatif dan data sebenarnya positif, sehingga di peroleh formulasi angka $20/26 \times 100 \% = 76,92 \%$

Tabel 14. Confusion Matrix

Aktual	Prediksi	
	Tepat	Telat
Tepat	(TP)12	(FN)1
Telat	(FP) 5	(TN)8

D. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, kesimpulan yang bisa diuraikan antara lain:

1. Metode Algoritma C.45 dapat digunakan untuk memprediksi keterlambatan santri dalam melakukan pembayaran SPP, dengan menggunakan variabel Penghasilan Orangtua, Tanggungan, Pekerjaan Ayah, Pembayaran Bulanan Pertama, Pembayaran Bulanan Kedua dan Pembayaran Bulanan Ketiga.
2. Telah dilakukan uji akurasi pada penerapan metode Algoritma C4.5 untuk prediksi keterlambatan santri dalam membayar SPP menggunakan confusion matrix dan diperoleh hasil akurasi sebesar 76,92%.

E. DAFTAR PUSTAKA

- [1] Arikunto, S. (2006). *Prosedur Penelitian Suatu Tindakan Praktik*. Jakarta: Rineka Cipta.
- [2] Eko Prasetyo. (2013). *Data Mining : Konsep Dan Aplikasi Menggunakan Matlab*. In *Journal of Chemical Information and Modeling*. <https://doi.org/10.1017/CBO9781107415324.004>
- [3] Eko Prasetyo. (2014). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta.
- [4] F. Gorunescu, *Data Mining Concept, Models and Techniques*. Verlag Berlin Heidelberg: Springer, 2011.
- [5] Hidayat, W., & Utami, A. (2022). Penerapan Metode Algoritma C4.5 Untuk Menentukan Kelayakan Calon Nasabah Pemegang Kartu Kredit Bank Mega Card center Kuningan. *TeknoIS : Jurnal Ilmiah Teknologi Informasi dan Sains*, 12(1), 31-48. doi:<https://doi.org/10.36350/jbs.v12i1.128>
- [6] Kusri and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009
- [7] Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. In *Discovering Knowledge in Data: An Introduction to Data Mining*.
- [8] Tri Herdiawan Apandi, Roby Bayu Maulana, Rian Piarna, Dwi Vernanda (2019). *Menganalisis Kemungkinan Keterlambatan Pembayaran SPP dengan ALGORITMA C4.5 (Studi Kasus POLITEKNIK TEDC BANDUNG)*
- [9] Turban, E., dkk, 2001, *Decision Support Systems and Intelligent Systems*, Andi Offset, Yogyakarta